IN THE CLAIMS:

Please cancel claims 3, 4, 13, 15, 16, 19, and 22-24 without prejudice.

Please amend claims 1, 2, 5-12, 14, 17, 18, 20, 21, 25, and 26 as follows:

1. (Currently Amended) A method for classifying electronically posted documents, the method comprising:

receiving a first document and a second document;

generating a first metadata summary ~~corresponding to~~ for said first document and a second metadata summary ~~corresponding to~~ for the second document, wherein the first metadata summary includes a first ~~summary sub-tree~~ plurality of sub-trees and the second metadata summary includes a second ~~sub-tree~~ plurality of sub-trees, and wherein ~~a sub-tree~~ each of the sub-trees includes a plurality of ~~list items~~ nodes;

comparing the first and second metadata summaries on a structural level by comparing ~~the list items~~ a structure of the sub-trees of the first metadata summary ~~sub-tree~~ with ~~the list items~~ a structure of the sub-trees of the second metadata summary ~~sub-tree~~; ~~and~~

identifying the first and second documents as distinct if the ~~list items~~ structures of the sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are not equivalent;

if the structures of the sub-trees of the first and second metadata summaries are equivalent, performing a further comparison of the first and second metadata summaries,

wherein the further comparison of the first and second metadata summaries includes the sub-steps of:

comparing the first and second metadata summaries on a textual level by comparing textual content from the first document that is contained in the sub-trees of the first metadata summary with textual content from the second document that is contained in the sub-trees of the second metadata summary; and

identifying the first and second documents as distinct if the textual content within the sub-trees of the first and second metadata summaries are not equivalent.

2. (Currently Amended) The method of claim 1, wherein ~~each list item includes at least one attribute having an attribute value,~~ the ~~method~~ further comparison of the first and second metadata summaries further ~~comprising~~ includes the sub-steps of:

before comparing the first and second metadata summaries on a textual level, comparing the first and second metadata summaries on an attribute level by comparing ~~the~~ attribute ~~value of a list item~~ values within the sub-trees of the first metadata summary ~~sub-tree~~ with ~~the~~ attribute ~~value of a list item~~ values within the sub-trees of the second metadata summary ~~sub-tree~~; and

identifying the first and second documents as distinct if the attribute values within the sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are not equivalent.

3-4. (Canceled)

5. (Currently Amended) The method of claim ~~4~~ 1, further comprising identifying the first and second documents as duplicates if the ~~text~~ textual content within the ~~list items~~ sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are equivalent.

6. (Currently Amended) The method of claim 5, further comprising removing the second metadata summary if ~~the structures of~~ the first and second ~~summary sub-trees~~ documents are ~~equivalent~~ identified as duplicates.

7. (Currently Amended) The method of claim 1, further comprising:

defining a first equivalence metadata table comprising:

a first row corresponding to the first metadata summary;

a second row corresponding to the second metadata summary;

a first column corresponding to the first metadata summary; and

a second column corresponding to the second metadata summary,

wherein the ~~process~~ step of identifying the first and second documents as distinct if the ~~list items~~ structures of the sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are not equivalent comprises storing a zero ~~binary~~ value in the first row and second column position of the first equivalence metadata table.

8. (Currently Amended) The method of claim 2, further comprising:

defining a first equivalence metadata table comprising:

a first row corresponding to the first metadata summary;

a second row corresponding to the second metadata summary;

a first column corresponding to the first metadata summary; and

a second column corresponding to the second metadata summary,

wherein the ~~process~~ step of identifying the first and second documents as distinct if the attribute values ~~of the list items~~ within the sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are not equivalent comprises storing a zero ~~binary~~ value in the first row and second column position of the first equivalence metadata table.

9. (Currently Amended) The method of claim ~~3~~ 7, ~~further comprising:~~

~~defining a first equivalence metadata table comprising:~~

~~a first row corresponding to the first metadata summary;~~

~~a second row corresponding to the second metadata summary;~~

~~a first column corresponding to the first metadata summary; and~~

~~a second column corresponding to the second metadata summary;~~

wherein the ~~process~~ step of identifying the first and second documents as distinct if the ~~text~~ textual content ~~of the list items~~ within the sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are not equivalent comprises ~~a~~ storing ~~of~~ a zero ~~binary~~ value in the first row and second column position of the first equivalence metadata table.

10. (Currently Amended) A method for classifying electronically posted documents, the method comprising:

receiving a plurality of documents;

generating a ~~respective plurality of~~ metadata ~~summaries corresponding to~~ summary for each of the ~~plurality of~~ received documents, wherein each of the metadata summaries includes a plurality of sub-trees and each of the sub-trees includes a plurality of nodes;

grouping a first subset of the ~~respective plurality of~~ metadata summaries into a first summary group, the first summary group ~~comprising~~ consisting of all of the metadata summaries having a first mime-type designation;

selecting a first metadata summary and a second metadata summary from the first summary group~~, wherein the first metadata summary includes a first summary sub-tree and the second metadata summary includes a second summary sub-tree and wherein a sub-tree includes a plurality of list items~~;

comparing the first and second metadata summaries on a structural level by comparing ~~the list items~~ a structure of the sub-trees of the first metadata summary ~~sub-tree~~ with ~~the list items~~ a structure of the sub-trees of the second metadata summary ~~sub-tree~~; and

identifying the first and second documents as distinct if the ~~list items~~ structures of the sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are not equivalent.

11. (Currently Amended) The method of claim 10, wherein the step of grouping further comprises grouping a second subset of the ~~respective~~ metadata summaries into a second summary group, the second summary group ~~comprising~~ consisting of all of the metadata summaries having a second mime-type designation.

12. (Currently Amended) A system for classifying electronically posted documents, the system comprising:

a metadata parser module coupled to receive electronically posted documents, the metadata parser configured to output ~~respective~~ a metadata ~~summaries~~ summary for each of the posted documents, wherein each ~~respective~~ of the metadata ~~summary~~ summaries comprises ~~one or more~~ a plurality of sub-trees~~, wherein a sub-tree~~ and each of the sub-trees includes a plurality of ~~list items and wherein a list item includes at least one attribute and at least one attribute value comprising text content~~ nodes;

a summary repository coupled to receive and store the ~~respective~~ metadata summaries; and

a summary consolidator coupled to the summary repository, the summary consolidator configured to:

compare ~~the list items of sub-trees,~~ a first of the metadata summaries and a second of the metadata summaries on a structural level by comparing a structure of the sub-trees of the first metadata summary with a structure of the sub-trees of the second metadata summary;

identify the first and second documents corresponding to the first and second metadata summaries as distinct if the ~~list items~~ structures of the sub-trees of the first and second metadata summaries are not equivalent; and~~, and delete duplicate metadata summaries from the summary repository~~

if the structures of the sub-trees of the first and second metadata summaries are equivalent, further compare the first and second metadata summaries,

wherein the further comparison of the first and second metadata summaries includes:

comparing the first and second metadata summaries on a textual level by comparing textual content from the first document that is contained in the sub-trees of the first metadata summary with textual content from the second document that is contained in the sub-trees of the second metadata summary; and

identifying the first and second documents corresponding to the first and second metadata summaries as distinct if the textual content within the sub-trees of the first and second metadata summaries are not equivalent.

13. (Canceled)

14. (Currently Amended) The system of claim ~~13~~ 12, wherein the ~~sub-tree comparator is configured to compare a metadata portion of the metadata summary~~ further comparison of the first and second metadata summaries further includes the:

before comparing the first and second metadata summaries on a textual level, comparing the first and second metadata summaries on an attribute level by comparing attribute values within the sub-trees of the first metadata summary with attribute values within the sub-trees of the second metadata summary; and

identifying the first and second documents corresponding to the first and second metadata summaries as distinct if the attribute values within the sub-trees of the first and second metadata summaries are not equivalent.

15. (Canceled)

16. (Canceled)

17. (Currently Amended) A program product for use in a computer system that executes program steps recorded in a computer-readable media to perform a method for classifying electronically posted documents, the program product comprising:

a record-able media;

a program of computer-readable instructions executable by the computer system to perform processes comprising the steps of:

receiving a first document and a second document;

generating a first metadata summary ~~corresponding to~~ for said first document and a second metadata summary ~~corresponding to~~ for the second document, wherein the first metadata summary includes a first ~~summary sub-tree~~ plurality of sub-trees and the second metadata summary includes a second ~~summary sub-tree~~ plurality of sub-trees, and wherein ~~a sub-tree~~ each of the sub-trees includes a plurality of ~~list items~~ nodes;

comparing the first and second metadata summaries on a structural level by comparing ~~the list items~~ a structure of the sub-trees of the first metadata summary ~~sub-tree~~ with ~~the list items~~ a structure of the sub-trees of the second metadata summary ~~sub-tree~~; ~~and~~

identifying the first and second documents as distinct if the ~~list items~~ structures of the sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are not equivalent;

if the structures of the sub-trees of the first and second metadata summaries are equivalent, performing a further comparison of the first and second metadata summaries,

wherein the further comparison of the first and second metadata summaries includes the sub-steps of:

comparing the first and second metadata summaries on a textual level by comparing textual content from the first document that is contained in the sub-trees of the first metadata summary with textual content from the second document that is contained in the sub-trees of the second metadata summary; and

identifying the first and second documents as distinct if the textual content within the sub-trees of the first and second metadata summaries are not equivalent.

18. (Currently Amended) The program product of claim 17, wherein ~~each list item includes at least one attribute having an attribute value,~~ the ~~program product method~~ further comparison of the first and second metadata summaries further ~~comprising the processes~~ includes the sub-steps of:

before comparing the first and second metadata summaries on a textual level, comparing the first and second metadata summaries on an attribute level by comparing ~~the~~ attribute ~~value of a list item~~ values within the sub-trees of the first metadata summary ~~sub-tree~~ with ~~the~~ attribute ~~value of a list item~~ values within the sub-trees of the second metadata summary ~~sub-tree~~; and

identifying the first and second documents as distinct if the attribute values within the sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are not equivalent.

19. (Canceled)

20. (Currently Amended) The program product of claim ~~19~~ 17, further comprising the ~~method~~ step of identifying the first and second documents as duplicates if the ~~text~~ textual content within the ~~list items~~ sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are equivalent.

21. (Currently Amended) The program product of claim 20, further comprising the ~~process~~ step of removing the second metadata summary if the first and second documents are identified as duplicates.

22-24. (Canceled)

25. (Currently Amended) A program product for use in a computer system that executes program steps recorded in a computer-readable media to perform a method for classifying electronically posted documents, the program product comprising:

a record-able media;

a program of computer-readable instructions executable by the computer system to perform method steps comprising:

receiving a plurality of documents;

generating a ~~respective plurality of~~ metadata ~~summaries corresponding to~~ summary for each of the ~~plurality of~~ received documents, wherein each of the metadata summaries includes a plurality of sub-trees and each of the sub-trees includes a plurality of nodes;

grouping a first subset of the ~~respective plurality of~~ metadata summaries ~~onto~~ into a first summary group, the first summary group ~~comprising~~ consisting of all of the metadata summaries having a first mime-type designation;

selecting a first metadata summary and a second metadata summary from the first summary group~~, wherein the first metadata summary includes a first summary sub-tree and the second metadata summary includes a second summary sub-tree and wherein a sub-tree includes a plurality of list items~~;

comparing the first and second metadata summaries on a structural level by comparing ~~the list items~~ a structure of the sub-trees of the first metadata summary ~~sub-tree~~ with ~~the list items~~ a structure of the sub-trees of the second metadata summary ~~sub-tree~~; and

identifying the first and second documents as distinct if the ~~list items~~ structures of the sub-trees of the first and second ~~summary sub-trees~~ metadata summaries are not equivalent.

26. (Currently Amended) The program product of claim 25, wherein the step of grouping further comprises grouping a second subset of the ~~respective~~ metadata summaries into a second summary group, the second summary group ~~comprising~~ consisting of all of the metadata summaries having a second mime-type designation.